

# Using Reed-Solomon codes in the $(U \mid U + V)$ construction and an application to cryptography

Irene Márquez-Corbella\*, Jean-Pierre Tillich†

February 1, 2016

## Abstract

In this paper we present a modification of Reed-Solomon codes that beats the Guruswami-Sudan  $1 - \sqrt{R}$  decoding radius of Reed-Solomon codes at low rates  $R$ . The idea is to choose Reed-Solomon codes  $U$  and  $V$  with appropriate rates in a  $(U \mid U + V)$  construction and to decode them with the Koetter-Vardy soft information decoder. We suggest to use a slightly more general version of these codes (but which has the same decoding performances as the  $(U \mid U + V)$ -construction) for code-based cryptography, namely to build a McEliece scheme. The point is here that these codes not only perform nearly as well (or even better in the low rate regime) as Reed-Solomon codes, their structure seems to avoid the Sidelnikov-Shestakov attack which broke a previous McEliece proposal based on generalized Reed-Solomon codes.

## 1 Introduction

**Improving upon the error correction performance of RS codes.** Reed-Solomon(RS) codes are among the most extensively used error correcting codes. It has long been known how to decode them up to half the minimum distance. This gives a decoding algorithm that corrects a fraction  $\frac{1-R}{2}$  of errors in an RS code of rate  $R$ . However, it is only in the late nineties that a breakthrough was obtained in this setting with Sudan's algorithm [19] and its improvement in [10] who showed how to go beyond this barrier with an algorithm which in its [10] version decodes any fraction of errors smaller than  $1 - \sqrt{R}$ . Later on, it was shown that this decoding algorithm could also be modified a little bit in order to cope with soft information on the errors [11]. Then it was realized in [16] that by a slight modification of RS codes and by an increase of the alphabet size it was possible to beat the  $1 - \sqrt{R}$  decoding radius. Their new family of codes is list decodable beyond this radius for low rate. Then, [9] improved on these codes by presenting a new family of codes, namely *folded RS codes* with a polynomial time decoding algorithm achieving the list decoding capacity  $1 - R - \epsilon$  for every rate  $R$  and  $\epsilon > 0$ .

The first purpose of this paper is to present another modification of RS codes that improves the fraction of errors that can be corrected. It consists in using RS codes in a  $(U \mid U + V)$  construction. We will show that, in the low rate regime, this class of codes outperforms rather significantly a classical RS code decoded with the Guruswami and Sudan algorithm [10]. The point is that this  $(U \mid U + V)$  code can be decoded in two steps :

1. First by subtracting the left part  $y_1$  to the right part  $y_2$  of the received vector  $(y_1 \mid y_2)$  and decoding it with respect to  $V$ . In such a case, we are left with decoding a RS code with about twice as many errors.

---

\*Inria, Email: irene.marquez-corbella@inria.fr.

†Inria, Email: jean-pierre.tillich@inria.fr.

2. Secondly, once we have recovered the right part  $v$  of the codeword, we can get a word  $(y_1 \mid y_2 - v)$  which should match two copies of a same word  $u$  of  $U$ . We can model this decoding problem by having some soft information.

It turns that the last channel error model is much less noisy than the original  $q$ -ary symmetric channel we started with. This soft information can be used in Koetter and Vardy's decoding algorithm. By this means we can choose  $U$  to be a RS code of much bigger rate than  $V$ . All in all, it turns out that by choosing  $U$  and  $V$  with appropriate rates we can beat the  $1 - \sqrt{R}$  bound in the low-rate regime.

It should be noted however that beating this  $1 - \sqrt{R}$  bound comes at the cost of having now an algorithm which does not work as for the aforementioned papers [19, 10, 16, 9] for every error of a given weight (the so called adversarial error model) but with probability  $1 - o(1)$  for errors of a given weight. However, contrarily to [16, 9] which results in a significant increase of the alphabet size of the code, our alphabet size actually decreases when compared to a RS code: it can be half of the code length and can be even smaller when we apply this construction recursively. Indeed, we will show that we can even improve the error correction performances by applying this construction again to the  $U$  and  $V$  components, i.e we can choose  $U$  to be a  $(U_1 \mid U_1 + V_1)$  code and we replace in the same way the RS code  $V$  by a  $(U_2 \mid U_2 + V_2)$  where  $U_1, U_2, V_1, V_2$  are RS codes.

**Application to cryptography.** In a second part of the paper we show how to use such codes (or codes derived by this approach) for cryptographic purposes, i.e. in a McEliece cryptosystem [12]. Recall that this public-key cryptosystem becomes more and more fashionable due to the threats on the most popular public key cryptosystems used today, namely RSA or DSA and ECDSA that would be completely broken by Shor's algorithm [17] if a large scale quantum computer could be built. Indeed, it is unlikely that a quantum computer would be able to threaten the security of the McEliece scheme because it is based on an NP-complete problem, namely decoding a linear code.

Probably one of the main drawback of McEliece when compared to RSA, DSA or ECDSA is its rather large key size. There have been several attempts to decrease the key size either by moving to more structured codes or to codes which have better error correction radius [14, 2]. Many of the structured algebraic proposals have been broken (see for instance [8]) but some of the quasi-cyclic code families that rely on modified LDPC codes or MDPC codes [1, 13] seem to resist cryptanalysis up to now. Relying on codes with better decoding performance met a similar fate, since here again many proposals of this kind have been broken. For instance [14] suggests to replace the binary Goppa codes of the original McEliece cryptosystem by Generalized RS codes (GRS) because of their much better decoding performance, but it got broken in [18].

There have been several attempts to repair GRS codes in this context either by adding random columns to the generator matrix of a GRS code [21] or by multiplying this generator matrix by the inverse of a sparse matrix with small average row weight  $m$  [2]. The [21] attempt got broken in [6] and the parameters of [2] got broken in [7] because  $m$  was chosen to be too small. The problem with the [2] approach is that the attack of [7] fails when  $m = 2$ , but the solution is then no more competitive when compared to a Goppa code because the decoding radius gets also scaled down by a multiplicative factor of  $m$  when compared to a GRS code.

We suggest here to revive the [2] approach with a generalized  $(U \mid U + V)$  scheme based on RS codes that has basically the same decoding capacity as a RS code and that looks in many respects like the [2] scheme with  $m = 2$ . This approach is also related to the approach pioneered by Wang in [20]. His code can be viewed as a certain subcode of our  $(U \mid U + V)$  construction. However, the decrease of the code rate results in a significant deterioration of the key size when compared to a code with the same error correction capacity as an RS code.

**Notation 1.** *Throughout the paper we will use the following notation.*

- A linear code of length  $n$ , dimension  $k$  and distance  $d$  over a finite field  $\mathbb{F}_q$  is referred to as an  $[n, k, d]_q$ -code. We denote its dual by  $\mathcal{C}^\perp$  which is defined to be an  $[n, n - k, d^\perp]_q$  code.
- The Hamming weight of a vector  $\mathbf{x}$ , denoted by  $w_H(\mathbf{x})$ , is defined as the number of nonzero elements.
- For a vector  $\mathbf{x}$  we denote by  $x(i)$  the  $i$ -th coordinate of  $\mathbf{x}$ .

## 2 $(U \mid U + V)$ construction

In this section, we recall a few facts about the  $(U \mid U + V)$  construction and their decoding.

**Definition 1.** Let  $U$  be an  $[n, k_u, d_u]_q$  code and  $V$  be an  $[n, k_v, d_v]_q$  code. We define the  $(U \mid U + V)$ -construction of  $U$  and  $V$  as the linear code:

$$\mathcal{C} = \{(\mathbf{u} \mid \mathbf{u} + \mathbf{v}) \mid \mathbf{u} \in U \text{ and } \mathbf{v} \in V\}.$$

The code  $\mathcal{C}$  has parameters  $[2n, k_u + k_v, \min\{2d_u, d_v\}]_q$ . A generator matrix of  $\mathcal{C}$  is:

$$\left( \begin{array}{c|c} G_u & G_u \\ \hline \mathbf{0} & G_v \end{array} \right) \in \mathbb{F}_q^{(k_u + k_v) \times 2n}$$

where  $G_u$  and  $G_v$  are generator matrices of  $U$  and  $V$  respectively.

### 2.1 Soft-decision decoding of $(U \mid U + V)$ codes

Let  $U$  and  $V$  be two codes with parameters  $[n, k_u, d_u]_q$  and  $[n, k_v, d_v]_q$ , respectively and  $\mathcal{C} \stackrel{\text{def}}{=} (U \mid U + V)$ . Suppose we transmit the codeword  $(\mathbf{u} \mid \mathbf{u} + \mathbf{v}) \in \mathcal{C}$  over a noisy channel and we receive the vector:  $\mathbf{y} = (\mathbf{y}_1 \mid \mathbf{y}_2) = (\mathbf{u} \mid \mathbf{u} + \mathbf{v}) + (\mathbf{e}_1 \mid \mathbf{e}_2)$ .

Decoding proceeds in two steps:

1. We combine  $\mathbf{y}_1$  and  $\mathbf{y}_2$  to find  $\mathbf{v}$ . That is, we decode  $\mathbf{y}_1 - \mathbf{y}_2 = \mathbf{v} + \mathbf{e}_2 - \mathbf{e}_1$  with respect to  $V$ . In the case of a soft decoder for  $V$  we compute first the probability

$$\text{prob}(v(i) = \alpha \mid y_1(i), y_2(i)) \quad \text{for all } \alpha \in \mathbb{F}_q.$$

2. We subtract  $(\mathbf{0} \mid \mathbf{v})$  to  $(\mathbf{y}_1 \mid \mathbf{y}_2)$  to get  $(\mathbf{u} + \mathbf{e}_1 \mid \mathbf{u} + \mathbf{e}_2) = (\mathbf{z}_1 \mid \mathbf{z}_2)$ . This is a noisy version of  $(\mathbf{u} \mid \mathbf{u})$ . We compute now for all  $\alpha \in \mathbb{F}_q$  and all coordinates  $i$  the probabilities  $\text{prob}(u(i) = \alpha \mid z_1(i), z_2(i))$  which is then passed to a soft decoder for  $U$ .

Let us explain how these probabilities can be computed. We assume that the noise model is given by a discrete memoryless channel with input alphabet  $\mathbb{F}_q$  and output alphabet  $\mathcal{Y}$ . The received vector is denoted by  $\mathbf{y} = (y(1), \dots, y(n)) \in \mathcal{Y}^n$  and the channel model specifies the transition probabilities with the following matrix  $\Pi_{\mathbf{y}}$

$$\Pi_{\mathbf{y}}^i(\alpha) = \text{prob}(\alpha \mid y(i)) \quad \text{for } i = 1, \dots, n \text{ and } \alpha \in \mathbb{F}_q.$$

$\Pi_{\mathbf{y}}^i$  denotes here the  $i$ -th column of  $\Pi_{\mathbf{y}}$  and  $\Pi_{\mathbf{y}}^i(\alpha)$  refers to the entry in the  $i$ -th column and row indexed by  $\alpha \in \mathbb{F}_q$ .

We will refer to  $\Pi$  as the  $q \times n$  **reliability matrix** of the codewords symbols. We will see below that this reliability matrix can also be obtained through the  $(U \mid U + V)$  decoding process. We will particularly be interested here in the  $q$ -ary symmetric channel model.

This channel is parametrized by the crossover probability  $p$  and the channel will be denoted here by  $q\text{-SC}_p$ . Here each time an element from  $\mathbb{F}_q$  is transmitted and is received either the

unchanged input symbol, with probability  $1 - p$ , or any of the other  $q - 1$  symbols, with probability  $\frac{p}{q-1}$ . In other words, the reliability matrix  $\Pi_{\mathbf{y}}$  for  $q$ -SC $_p$  is defined as follows:

$$\Pi_{\mathbf{y}}^i(\alpha) = \text{prob}(\alpha | y(i)) = \begin{cases} 1 - p & \text{if } \alpha = y(i) \\ \frac{p}{q-1} & \text{if } \alpha \neq y(i) \end{cases}$$

Thus, all columns of  $\Pi_{\mathbf{y}}$  are identical up to permutation:

$$\Pi_{\mathbf{y}}^i = \begin{pmatrix} 1 - p \\ \frac{p}{q-1} \\ \vdots \\ \frac{p}{q-1} \end{pmatrix} \text{ (up to permutation)}$$

with  $i = 1, \dots, n$ .

Let us recall now how the reliability matrices for the decoder of  $U$  and  $V$  are computed from the initial reliability matrix.

**Reliability matrix for the  $V$ -decoder** We call in what follows the error model for the  $V$ -decoder the **sum model** and denote the associated reliability matrix by  $\Pi \oplus \Pi$  when  $\Pi$  is the initial reliability matrix. Recall that before decoding, for each symbol  $X$  of  $V$  that we want to decode we subtract two symbols  $X_1$  and  $X_2$  of the  $(U | U + V)$  code:

$$X = X_2 - X_1.$$

For each of these symbols we have a reliability information  $\text{prob}(X_1 = \alpha | Y_1)$  and  $\text{prob}(X_2 = \beta | Y_2)$  where  $Y_1$  and  $Y_2$  are random variables that are initially the received symbols corresponding to  $X_1$  and  $X_2$  after transmission on the noisy channel but that become sets of received symbols when we iterate the  $(U | U + V)$  construction as will be seen. When  $X_1$  and  $X_2$  are uniformly distributed it can be verified that

$$\text{prob}(X = \alpha | Y_1, Y_2) = \sum_{\beta \in \mathbb{F}_q} \text{prob}(X_1 = \beta | Y_1) \cdot \text{prob}(Y_2 = \alpha + \beta | Y_2)$$

This leads to the following definition.

$$(\Pi \oplus \Pi)_{\mathbf{y}}^i(\alpha) \stackrel{\text{def}}{=} \sum_{\beta \in \mathbb{F}_q} \Pi_{\mathbf{y}_1}^i(\beta) \cdot \Pi_{\mathbf{y}_2}^i(\alpha + \beta)$$

where  $\mathbf{y}_1$  and  $\mathbf{y}_2$  are the realizations of the channel transmission of  $u$  and  $u + v$  respectively. We also denote by  $\Pi_{\mathbf{y}_1}^i \oplus \Pi_{\mathbf{y}_2}^i$ , where each element represents a column vector, the  $i$ -th column of the  $\Pi \oplus \Pi$  matrix.

**Reliability matrix for the  $U$ -decoder** The computation of  $\text{prob}(u(i) = \alpha | z_1(i), z_2(i))$  can be performed by computing the probability that a uniformly distributed random variable over  $\mathbb{F}_q$  is equal to  $\alpha$  given two received symbols  $y_1$  and  $y_2$  for  $X$  sent over two memoryless channels (and which are chosen uniformly at random in  $\mathbb{F}_q$ ). This probability is readily seen to be equal to

$$\text{prob}(X = \alpha | y_1 \text{ and } y_2) = \frac{\text{prob}(X = \alpha | y_1) \cdot \text{prob}(X = \alpha | y_2)}{\sum_{\beta \in \mathbb{F}_q} \text{prob}(X = \beta | y_1) \cdot \text{prob}(X = \beta | y_2)}$$

We denote by  $\Pi \times \Pi$  the reliability matrix (the input) to a soft-decision decoding algorithm for the code  $U$ . Thus, each element of the reliability matrix  $\Pi \times \Pi$  related to the aforementioned quantities  $\mathbf{y}$  and  $\mathbf{v}$  is defined by:

$$(\Pi \times \Pi)_{\mathbf{y}, \mathbf{v}}^i(\alpha) = \frac{\Pi_{\mathbf{y}_1}^i(\alpha) \cdot \Pi_{\mathbf{y}_2}^i(\alpha + v(i))}{\sum_{\beta \in \mathbb{F}_q} \Pi_{\mathbf{y}_1}^i(\beta) \cdot \Pi_{\mathbf{y}_2}^i(\beta + v(i))}.$$

To simplify notation we will generally avoid the dependency on  $\mathbf{v}$  and simply write  $(\Pi \times \Pi)_{\mathbf{y}}$ .

## 2.2 Algebraic-soft decision decoding of RS codes

Let us recall how the Koetter-Vardy soft decoder [11] can be analyzed. By [11, Theorem 12] their decoding algorithm outputs a list that contains the codeword  $\mathbf{c} \in C$  if

$$\frac{\langle \Pi, \lfloor \mathbf{c} \rfloor \rangle}{\sqrt{\langle \Pi, \Pi \rangle}} \geq \sqrt{k-1} + o(1)$$

as the codelength  $n$  tends to infinity, where  $\lfloor \mathbf{c} \rfloor$  represents a  $q \times n$  matrix with entries  $c_{i,\alpha} = 1$  if  $c_i = \alpha$ , and 0 otherwise; and  $\langle A, B \rangle$  denotes the inner product of the two  $q \times n$  matrices  $A$  and  $B$ , i.e.

$$\langle A, B \rangle \stackrel{\text{def}}{=} \sum_{i=1}^q \sum_{j=1}^n a_{i,j} b_{i,j}.$$

The algorithm uses a parameter  $s$  (the total number of interpolation points counted with multiplicity). The Little-O  $o(1)$  depends on the choice of this parameter and the parameters  $n$  and  $q$ . It can be chosen as a function of  $q$  in such a way that, when  $\langle \Pi, \Pi \rangle$  has a lower bound given by some positive constant then, the Little-O of this formula is bounded from above by a function of  $q$  that goes to 0 as  $q$  goes to infinity. We will consider here only discrete symmetric channel models that are defined below. Let us first introduce some notation.

**Notation 2** (Probability error vector of a Discrete Memoryless Channel (DMC)). *For a given DMC with  $q$ -ary inputs we denote by  $\pi$  the probability vector  $\pi = (\text{prob}(x = \alpha|y))_{\alpha \in \mathbb{F}_q}$  where  $x$  is the symbol that has been sent through the channel and  $y$  is the received symbol. For a vector  $\mathbf{x} = (x(\beta))_{\beta \in \mathbb{F}_q}$  we denote by  $\mathbf{x}^{+\alpha}$  the vector  $\mathbf{x}^{+\alpha} = (x(\beta + \alpha))_{\beta \in \mathbb{F}_q}$ .*

By viewing  $\pi$  as a random variable (namely as a function of the random variable  $y$ ), we define as in [3] a symmetric channel by

**Definition 2** (discrete symmetric channel with  $q$ -ary inputs). A DMC with  $q$ -ary inputs is said to be symmetric if and only if for any  $\alpha$  in  $\mathbb{F}_q$  we have

$$p(\alpha) \text{prob}(\pi = \mathbf{p}) = p(0) \text{prob}(\pi = \mathbf{p}^{+\alpha}). \quad (1)$$

Note that this implies that in a discrete symmetric channel, for any possible realization  $\mathbf{p}$  of the probability vector  $\pi$  (i.e. when  $\text{prob}(\pi = \mathbf{p}) \neq 0$ ) we necessarily have  $p(0) \neq 0$ , since otherwise we would have  $p(\alpha) = 0$  for all  $\alpha \neq 0$ , a contradiction with the fact that  $\mathbf{p}$  is a probability vector  $\sum_{\alpha \in \mathbb{F}_q} p(\alpha) = 1$ . It is proved in [3] that symmetric channels are closed under the  $\oplus$  and  $\times$  operations on channels defined in Subsection 2.1. We give now the asymptotic behavior for a symmetric channel of the Koetter-Vardy decoder, but before doing this we will need a few lemmas.

**Lemma 3.** *Let  $\pi = (\pi(\alpha))_{\alpha \in \mathbb{F}_q}$  be the probability vector associated to a discrete symmetric channel with  $q$ -ary inputs. Then*

$$\mathbb{E}(\pi(0)) = \mathbb{E}(\|\pi\|^2), \quad \text{with } \|\pi\|^2 \stackrel{\text{def}}{=} \sum_{\alpha \in \mathbb{F}_q} \pi(\alpha)^2.$$

*Proof.* Let us begin by observing that for a symmetric channel if  $\mathbf{p}$  is a possible realization of the probability vector  $\pi$ , then  $\mathbf{p}^{+\alpha}$  is also a possible realization of this probability vector as soon as  $p(\alpha) \neq 0$ . This motivates to introduce the equivalence relation between probability vectors over  $\mathbb{F}_q$  :

$$\mathbf{p} \equiv \mathbf{q} \text{ iff there exists } \alpha \in \mathbb{F}_q \text{ for which } \mathbf{p}^{+\alpha} = \mathbf{q}.$$

We consider for our DMC the set of all equivalence classes of the probability vector  $\pi$  and denote by  $\mathcal{R}$  a set of representatives of such equivalence classes. For a representative  $\mathbf{r}$  of an

equivalence class we denote by  $c(\mathbf{r})$  the class to which it belongs. Observe now that for such a representative  $\mathbf{r} = (r(\beta))_{\beta \in \mathbb{F}_q}$  we have

$$\text{prob}(\pi = \mathbf{r}^{+\alpha} | \pi \in c(\mathbf{r})) = \sum_{\alpha \in \mathbb{F}_q} \frac{Kr(\alpha)}{r(0)} \quad (2)$$

for some constant  $K > 0$  by using (1). Since

$$\sum_{\alpha \in \mathbb{F}_q} \text{prob}(\pi = \mathbf{r}^{+\alpha} | \pi \in c(\mathbf{r})) = 1 \quad \text{and} \quad \sum_{\alpha \in \mathbb{F}_q} r(\alpha) = 1$$

we necessarily have  $K = r(0)$ . Therefore,

$$\sum_{\alpha \in \mathbb{F}_q} r^{+\alpha}(0) \text{prob}(\pi = \mathbf{r}^{+\alpha} | \pi \in c(\mathbf{r})) = \sum_{\alpha \in \mathbb{F}_q} r(\alpha) \frac{Kr(\alpha)}{r(0)} = \sum_{\alpha \in \mathbb{F}_q} r(\alpha)^2.$$

This implies that

$$\begin{aligned} \mathbb{E}(\pi(0)) &= \sum_{\mathbf{r} \in \mathcal{R}} \text{prob}(\pi \in c(\mathbf{r})) \sum_{\alpha \in \mathbb{F}_q} r^{+\alpha}(0) \text{prob}(\pi = \mathbf{r}^{+\alpha} | \pi \in c(\mathbf{r})) \\ &= \sum_{\mathbf{r} \in \mathcal{R}} \text{prob}(\pi \in c(\mathbf{r})) \sum_{\alpha \in \mathbb{F}_q} r(\alpha)^2 \\ &= \sum_{\mathbf{r} \in \mathcal{R}} \text{prob}(\pi \in c(\mathbf{r})) \|\mathbf{r}\|^2 = \mathbb{E}(\|\pi\|^2). \end{aligned}$$

□

Let us recall the Chebyshev inequality which says that for any random variable  $X$  we have

$$\text{prob}(|X - \mathbb{E}(X)| \geq t) \leq \frac{\text{Var}(X)}{t^2} \quad (3)$$

We are going to use this result with  $X = \langle \Pi, [\mathbf{0}] \rangle$  and  $X = \langle \Pi, \Pi \rangle$ . This leads to the following concentration results.

**Lemma 4.** *Let  $\epsilon > 0$ . We have*

$$\text{prob}(\langle \Pi, [\mathbf{0}] \rangle \leq (1 - \epsilon)n\mathbb{E}(\|\pi\|^2)) \leq \frac{1}{n\epsilon^2 (\mathbb{E}(\|\pi\|^2))^2} \quad (4)$$

$$\text{prob}(\langle \Pi, \Pi \rangle \geq (1 + \epsilon)n\mathbb{E}(\|\pi\|^2)) \leq \frac{1}{n\epsilon^2 (\mathbb{E}(\|\pi\|^2))^2} \quad (5)$$

*Proof.* Let us first prove (4). First observe that

$$\langle \Pi, [\mathbf{0}] \rangle = \sum_{i=1}^n \Pi^i(0)$$

By linearity of expectation and Lemma 3 we have

$$\mathbb{E}\{\langle \Pi, [\mathbf{0}] \rangle\} = n\mathbb{E}(\pi(0)) = n\mathbb{E}(\|\pi\|^2). \quad (6)$$

Since the column vectors  $\Pi^1(0), \Pi^2(0), \dots, \Pi^n(0)$  are independent random variables we also obtain

$$\text{Var}\{\langle \Pi, [\mathbf{0}] \rangle\} = \text{Var}\left(\sum_{i=1}^n \Pi^i(0)\right) = n\text{Var}(\pi(0)) \leq n. \quad (7)$$

From this we deduce

$$\begin{aligned}
\text{prob} \left( \langle \Pi, [\mathbf{0}] \rangle \leq (1 - \epsilon)n\mathbb{E}(\|\pi\|^2) \right) &\leq \text{prob} \left( \left| \langle \Pi, [\mathbf{0}] \rangle - n\mathbb{E}(\|\pi\|^2) \right| \geq \epsilon n\mathbb{E}(\|\pi\|^2) \right) \\
&\leq \frac{\text{Var} \{ \langle \Pi, [\mathbf{0}] \rangle \}}{\epsilon^2 n^2 \left( \mathbb{E}(\|\pi\|^2) \right)^2} \text{ by (3) and (6)} \\
&\leq \frac{1}{\epsilon^2 n \left( \mathbb{E}(\|\pi\|^2) \right)^2} \text{ by (7)}
\end{aligned}$$

This proves (4). The second statement follows by similar considerations. We have in this case

$$\mathbb{E} \{ \langle \Pi, \Pi \rangle \} = n\mathbb{E} \left( \|\pi\|^2 \right) \quad (8)$$

$$\text{Var} \{ \langle \Pi, \Pi \rangle \} \leq n \quad (9)$$

This can be used to prove that

$$\begin{aligned}
\text{prob} \left( \langle \Pi, \Pi \rangle \geq (1 + \epsilon)n\mathbb{E}(\|\pi\|^2) \right) &\leq \text{prob} \left( \left| \langle \Pi, \Pi \rangle - n\mathbb{E}(\|\pi\|^2) \right| \geq \epsilon n\mathbb{E}(\|\pi\|^2) \right) \\
&\leq \frac{\text{Var} \{ \langle \Pi, \Pi \rangle \}}{\epsilon^2 n^2 \left( \mathbb{E}(\|\pi\|^2) \right)^2} \text{ by (3) and (8)} \\
&\leq \frac{1}{\epsilon^2 n \left( \mathbb{E}(\|\pi\|^2) \right)^2} \text{ by (9)}
\end{aligned}$$

□

We are ready now to prove the following theorem which gives a (tight) lower bound on the error-correction capacity of the Koetter-Vardy decoding algorithm over a discrete memoryless channel.

**Theorem 5.** *Let  $(\mathcal{C}_n)_{n \geq 1}$  be an infinite family of Reed-Solomon codes of rate  $\leq R$ . Denote by  $q_n$  the alphabet size of  $\mathcal{C}_n$  that is assumed to be a non decreasing sequence that goes to infinity with  $n$ . Consider an infinite family of  $q_n$ -ary symmetric channels with associated probability error vectors  $\pi_n$  such that  $\mathbb{E}(\|\pi_n\|^2)$  has a limit as  $n$  tends to infinity. Let*

$$C_{KV} \stackrel{\text{def}}{=} \lim_{n \rightarrow \infty} \mathbb{E}(\|\pi_n\|^2).$$

*This infinite family of codes can be decoded correctly by the Koetter-Vardy decoding algorithm with probability  $1 - o(1)$  as  $n$  tends to infinity as soon as there exists  $\epsilon > 0$  such that*

$$R \leq C_{KV} - \epsilon.$$

*Remark 1.* Let us observe that for the  $q$ -SC <sub>$p$</sub>  we have

$$\mathbb{E}(\|\pi\|^2) = (1 - p)^2 + (q - 1) \frac{p^2}{(q - 1)^2} = (1 - p)^2 + \mathcal{O}\left(\frac{1}{q}\right).$$

By letting  $q$  going to infinity, we recover in this way the performance of the Guruswami-Sudan algorithm which works as soon as  $R < (1 - p)^2$ .

*Proof of Theorem 5.* Without loss of generality we may assume that the codeword that was sent is the zero codeword  $\mathbf{0}$ . Let  $n_0$  be such that

$$C_{KV} - \frac{\epsilon}{2} \leq \mathbb{E}(\|\pi_n\|^2) \leq C_{KV} + \frac{\epsilon}{2} \quad (10)$$

for any  $n \geq n_0$ . For  $n \geq n_0$  we can write that the rate  $R_n$  of  $\mathcal{C}_n$  satisfies

$$\begin{aligned} R_n &\leq C_{KV} - \epsilon \\ &\leq \mathbb{E}(\|\pi_n\|^2) - \epsilon/2 \end{aligned} \quad (11)$$

Let  $K$  and  $N$  be the dimension and the length of  $\mathcal{C}_n$ . In such a case we have

$$\begin{aligned} \sqrt{\frac{K-1}{N}} &\leq \sqrt{\frac{K}{N}} \\ &\leq \sqrt{\mathbb{E}(\|\pi_n\|^2) - \epsilon/2} \quad (\text{by (11)}) \\ &\leq \sqrt{\mathbb{E}(\|\pi_n\|^2)} \sqrt{1 - \frac{\epsilon}{2\mathbb{E}(\|\pi_n\|^2)}} \\ &\leq \sqrt{\mathbb{E}(\|\pi_n\|^2)} \left(1 - \frac{\epsilon}{4\mathbb{E}(\|\pi_n\|^2)}\right) \quad (\text{since } \sqrt{1-x} \leq 1 - \frac{x}{2}) \\ &\leq \sqrt{\mathbb{E}(\|\pi_n\|^2)} \left(1 - \frac{\epsilon}{4C_{KV} + 2\epsilon}\right) \quad (\text{by (10)}) \end{aligned} \quad (12)$$

Let  $\delta$  be a positive constant that we are going to choose afterwards. Note that if an event  $\mathcal{E}_1$  has probability  $\geq 1 - \epsilon_1$  and another event  $\mathcal{E}_2$  has probability  $\geq 1 - \epsilon_2$ , then

$$\text{prob}(\mathcal{E}_1 \cap \mathcal{E}_2) = \text{prob}(\mathcal{E}_1) + \text{prob}(\mathcal{E}_2) - \text{prob}(\mathcal{E}_1 \cup \mathcal{E}_2) \geq 1 - \epsilon_1 + 1 - \epsilon_2 - 1 = 1 - \epsilon_1 - \epsilon_2.$$

We can use this remark together with Lemma 4 to deduce that with probability greater than or equal to  $1 - \frac{2}{N\delta^2(\mathbb{E}(\|\pi_n\|^2))^2}$  we have at the same time

$$\langle \Pi_n, [\mathbf{0}] \rangle \geq (1 - \delta)N\mathbb{E}(\|\pi_n\|^2) \quad (13)$$

$$\langle \Pi_n, \Pi_n \rangle \leq (1 + \delta)N\mathbb{E}(\|\pi_n\|^2) \quad (14)$$

In such a case we have

$$\frac{\langle \Pi_n, [\mathbf{0}] \rangle}{\sqrt{\langle \Pi_n, \Pi_n \rangle}} \geq \frac{1 - \delta}{\sqrt{1 + \delta}} \sqrt{N} \sqrt{\mathbb{E}(\|\pi_n\|^2)} \quad (15)$$

There exists  $x_0 > 0$  such that for every  $x \in [0, x_0]$  we have

$$\frac{1 - x}{\sqrt{1 + x}} \leq 1 - 2x.$$

Therefore for  $\delta \leq x_0$  we have in the aforementioned case

$$\frac{\langle \Pi_n, [\mathbf{0}] \rangle}{\sqrt{\langle \Pi_n, \Pi_n \rangle}} \geq (1 - 2\delta) \sqrt{N} \sqrt{\mathbb{E}(\|\pi_n\|^2)} \quad (16)$$

Let us choose now  $\delta$  such that

$$\delta = \min\left(x_0, \frac{\epsilon'}{4}\right).$$

where  $\epsilon' \stackrel{\text{def}}{=} \frac{\epsilon}{4C_{KV} + 2\epsilon}$ . This choice implies

$$\begin{aligned} \frac{\langle \Pi_n, [\mathbf{0}] \rangle}{\sqrt{\langle \Pi_n, \Pi_n \rangle}} &\geq (1 - \epsilon'/2) \sqrt{N} \sqrt{\mathbb{E}(\|\pi_n\|^2)} \\ &\geq \frac{1 - \epsilon'/2}{1 - \epsilon'} \sqrt{\frac{K-1}{N}} \end{aligned} \quad (17)$$



where we used (12) for the last inequality. Therefore we deduce that in the aforementioned case (i.e. when (13) and (14) both hold), that we can choose  $s$  appropriately in the Koetter-Vardy algorithm so that the codeword  $\mathbf{0}$  is in the list output by the algorithm. The probability that (17) is satisfied is greater than or equal to  $1 - \frac{2}{N\delta^2(\mathbb{E}(\|\pi_n\|^2))^2}$  which is also greater than or equal to (by using (10))  $1 - \frac{2}{N\delta^2(C_{KV} - \frac{\epsilon}{2})^2}$  which goes to 1 as  $N$  goes to infinity.  $\square$

### 3 Correcting errors beyond the Guruswami-Sudan bound

#### 3.1 The $(U | U + V)$ -construction

Now suppose we choose  $U$  and  $V$  as RS codes in a  $(U | U + V)$  construction. We start with a  $q$ -ary symmetric channel with error probability  $p$ . Recall that the reliability matrix for the  $U$ -decoder is  $\Pi_1 = \Pi \times \Pi$  whereas for the  $V$ -decoder it is  $\Pi_2 = \Pi \oplus \Pi$ .

**Lemma 6.** *Let  $\pi_U$  and  $\pi_V$  be the probability vectors corresponding to decoding the codes  $U$  and  $V$  respectively.*

- *The channel error model of the code  $V$  is a  $q$ -SC $_{p'}$  with  $p' = 2p - p^2$  and*

$$\mathbb{E}(\|\pi_V\|^2) = (1 - p)^4 + \mathcal{O}\left(\frac{1}{q}\right).$$

- *For the channel error model of the code  $U$  we have*

$$\mathbb{E}(\|\pi_U\|^2) = \frac{(p + 2)(p - 1)^2}{2 - p} + \mathcal{O}\left(\frac{1}{q}\right).$$

*Proof.* The proof of this Lemma can be found in Appendix A  $\square$

**Proposition 7.** *As  $q$  tends to infinity, the  $(U | U + V)$ -construction can be decoded correctly by the Koetter-Vardy decoding algorithm with probability  $1 - o(1)$  if*

$$R < \frac{(p^3 - 4p^2 + 4p - 4)(1 - p)^2}{2(p - 2)}$$

*Proof.* The  $(U | U + V)$ -construction can be decoded correctly by the Koetter-Vardy decoding algorithm if it decodes correctly  $U$  and  $V$ . By Theorem 5 decoding succeeds with probability  $1 - o(1)$  when we choose the rate  $R_U$  of  $U$  to be any positive number below  $\mathbb{E}(\|\pi_U\|^2)$  and the rate of  $V$  any positive number below  $\mathbb{E}(\|\pi_V\|^2)$ . Since the rate  $R$  of the  $(U | U + V)$  construction is equal to  $\frac{R_U + R_V}{2}$  decoding succeeds if

$$R < \lim_{q \rightarrow \infty} \frac{\mathbb{E}\{\|\pi_U\|^2\} + \mathbb{E}\{\|\pi_V\|^2\}}{2} = \frac{(p^3 - 4p^2 + 4p - 4)(1 - p)^2}{2(p - 2)}.$$

$\square$

From Figure 2 we deduce that the  $(U | U + V)$  decoder outperforms the RS decoder with Guruswami-Sudan as soon as  $R < 0.168$ .

### 3.2 Recursive application of the $(U | U + V)$ construction

Now we will study what happens over the  $q$ -SC $_p$  if we apply recursively the  $(U | U + V)$  construction. So we start with a  $(U | U + V)$  code, we choose  $U$  to be a  $(U_1 | U_1 + V_1)$  code and  $V$  to be a  $(U_2 | U_2 + V_2)$  code, where  $U_1, U_2, V_1$  and  $V_2$  are RS codes over the same alphabet  $\mathbb{F}_q$  and of the same length. In other words, we look for a code of the form

$$(U_1 | U_1 + V_1 | U_1 + U_2 | U_1 + U_2 + V_1 + V_2) = \{(\mathbf{u}_1 | \mathbf{u}_1 + \mathbf{v}_1 | \mathbf{u}_1 + \mathbf{u}_2 | \mathbf{u}_1 + \mathbf{u}_2 + \mathbf{v}_1 + \mathbf{v}_2) : \mathbf{u}_i \in U_i, \mathbf{v}_i \in V_i\}$$

From Lemma 6 we obtain the channel error models for decoding  $U_1, V_1, U_2$  and  $V_2$  respectively, their reliability matrices are given by  $\Pi_1 \times \Pi_1, \Pi_1 \oplus \Pi_1, \Pi_2 \times \Pi_2$  and  $\Pi_2 \oplus \Pi_2$  respectively (see Fig. 3). We let  $p' \stackrel{\text{def}}{=} 2p - p^2$ .

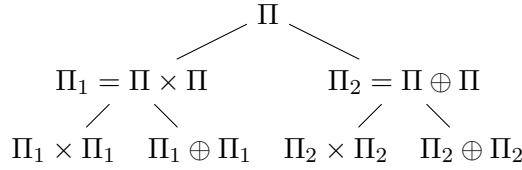


Fig. 1: The channel error models for  $(U_1 | U_1 + V_1 | U_1 + U_2 | U_1 + U_2 + V_1 + V_2)$ .

**Lemma 8.** Let  $\pi_{U_i}$  and  $\pi_{V_i}$  be the probability vectors corresponding to decoding the codes  $U_i$ 's and  $V_i$ 's.

- The channel error model of the code  $V_2$  is a  $q$ -SC $_{p''}$  with  $p'' = 2p' - p'^2$  and

$$\mathbb{E}(\|\pi_{V_2}\|^2) = (1 - p)^8 + \mathcal{O}\left(\frac{1}{q}\right);$$

- $\mathbb{E}(\|\pi_{U_2}\|^2) = \frac{(2+p')(1-p')^2}{(2-p')^2} + \mathcal{O}\left(\frac{1}{q}\right);$
- $\mathbb{E}(\|\pi_{V_1}\|^2) = (1 - p)^4 \left( \frac{2+3p+8p^2-4p^3}{2-p} \right) + \mathcal{O}\left(\frac{1}{q}\right);$
- $\mathbb{E}(\|\pi_{U_1}\|^2) = \frac{(5p^3-6p^2-5p-4)(1-p)^2}{4-3p} + \mathcal{O}\left(\frac{1}{q}\right).$

*Proof.* The proof of this Lemma can be found in Appendix B □

**Proposition 9.** As  $q$  tends to infinity, the  $(U_1 | U_1 + V_1 | U_1 + U_2 | U_1 + U_2 + V_1 + V_2)$ -construction can be decoded correctly by the Koetter-Vardy decoding algorithm with probability  $1 - o(1)$  if

$$R < \frac{(3p^{10} - 34p^9 + 187p^8 - 628p^7 + 1376p^6 - 2016p^5 + 1970p^4 - 1272p^3 + 568p^2 - 208p + 64)(p - 1)^2}{4(p^2 - 2p + 2)(3p - 4)(p - 2)}$$

*Proof.* Decoding of  $(U_1 | U_1 + V_1 | U_1 + U_2 | U_1 + U_2 + V_1 + V_2)$  succeeds if the Koetter-Vardy decoder is able to decode correctly  $U_1, U_2, V_1$  and  $V_2$ . This happens with probability  $1 - o(1)$  as soon as the rates  $R_{U_1}, R_{U_2}, R_{V_1}$  and  $R_{V_2}$  of these codes satisfy for some  $\epsilon > 0$

$$\begin{aligned} R_{U_1} &\leq \mathbb{E}(\|\pi_{U_1}\|^2) - \epsilon \\ R_{U_2} &\leq \mathbb{E}(\|\pi_{U_2}\|^2) - \epsilon \\ R_{V_1} &\leq \mathbb{E}(\|\pi_{V_1}\|^2) - \epsilon \\ R_{V_2} &\leq \mathbb{E}(\|\pi_{V_2}\|^2) - \epsilon \end{aligned}$$

Since the rate  $R$  of  $(U_1 | U_1 + V_1 | U_1 + U_2 | U_1 + U_2 + V_1 + V_2)$  is given by

$$R = \frac{R_{U_1} + R_{U_2} + R_{V_1} + R_{V_2}}{4}$$

we finally obtain that decoding succeeds with probability  $1 - o(1)$  as soon as the rate  $R$  is chosen such that

$$R < \lim_{q \rightarrow \infty} \frac{\sum_{i=1}^2 \mathbb{E} \left\{ \|\pi_{U_i}\|^2 \right\} + \mathbb{E} \left\{ \|\pi_{V_i}\|^2 \right\}}{4}$$

This implies the proposition by plugging the value of these expectations by using Lemma 8.  $\square$

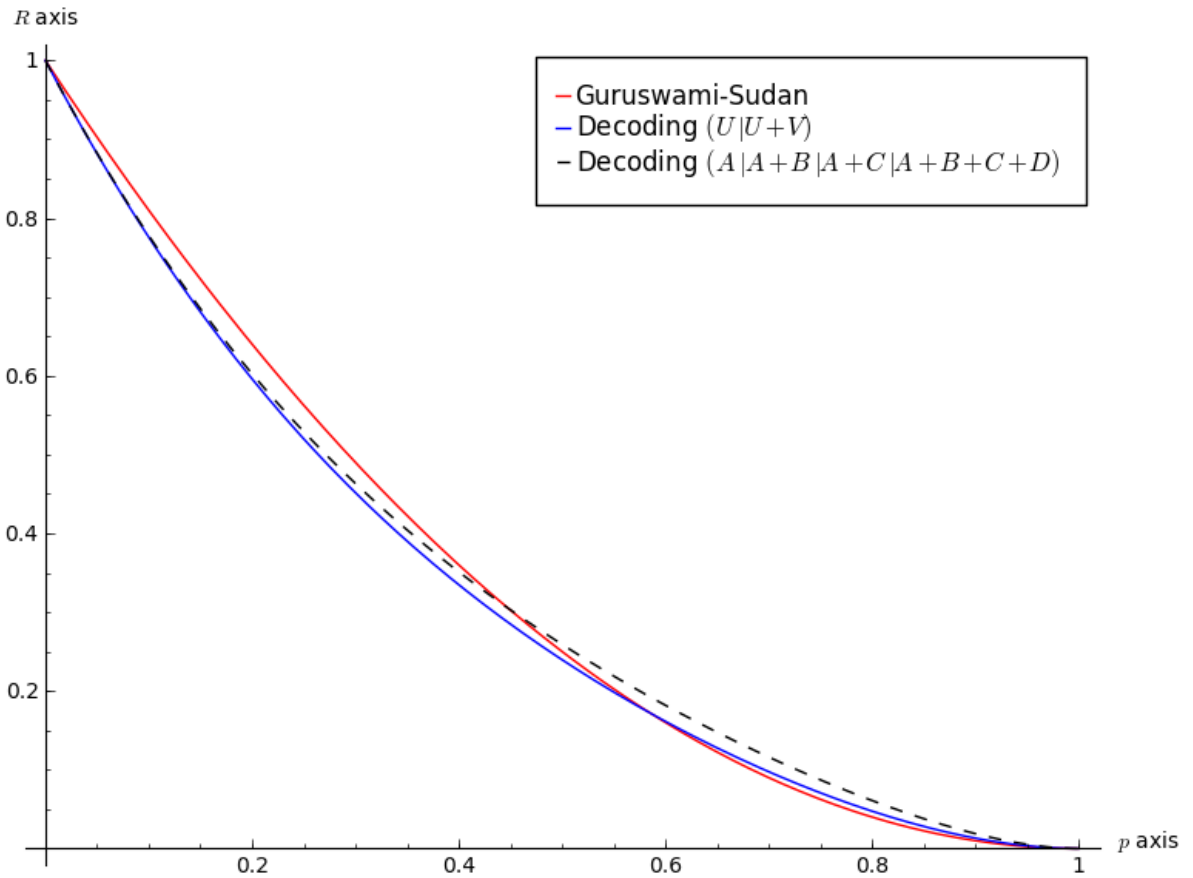


Fig. 2: Rate plotted against the crossover error probability  $p$  for several algorithms. The red line refers to the Guruswami-Sudan algorithm, the blue line to the  $(U | U + V)$ -construction and the dashed line to the  $(U_1 | U_1 + V_1 | U_1 + U_2 | U_1 + U_2 + V_1 + V_2)$ -construction.

From Figure 2 we deduce that if we apply twice the  $(U | U + V)$ -construction we get better performance than decoding a classical RS code with the Guruswami-Sudan decoder for low rate codes, specifically for  $R < 0.3$ .

## 4 A new Mc-Eliece scheme

As we have seen, this  $(U | U + V)$  construction gives codes which in the low rate regime have even better error correction capacities than a standard RS code. This suggests to use such codes

in a McEliece cryptosystem to replace the original Goppa codes. These  $(U | U + V)$  codes do not only have a better error correction capacity, they also allow to avoid the Sidelnikov-Shestakov attack [18] that broke a previous proposal based on GRS codes [14]. Furthermore we can even strengthen the security of this scheme by using instead of the  $(U | U + V)$  construction a generalized  $(U | U + V)$  code which has trivially the same error-correction capacity as the  $(U | U + V)$  construction but with better minimum distance properties which seems essential to avoid attacks based on finding minimum weight codewords in the code and trying to unravel the code structure from those minimum weight codewords. Analyzing precisely attacks of this kind needs however additional tools due to the peculiar structure of these generalized  $(U | U + V)$  codes (it is for instance inappropriate to use the analysis done for random codes) and is out of scope of this paper.

**Definition 3.** Let  $(U, V)$  be a pair of codes with parameters  $[n, k_u, d_u]_q$  and  $[n, k_v, d_v]_q$ , respectively. Consider the following matrix

$$\mathbf{D} = \left( \begin{array}{c|c} \mathbf{D}_1 & \mathbf{D}_3 \\ \hline \mathbf{D}_2 & \mathbf{D}_4 \end{array} \right) \in \mathbb{F}_q^{n \times n}$$

where the  $\mathbf{D}_i$ 's are diagonal matrices such that  $\mathbf{D}$  is non singular. We define the generalized  $(U | U + V)$ -construction of  $U$  and  $V$  with respect to  $\mathbf{D}$  as the matrix product code:

$$\{(\mathbf{u}\mathbf{D}_1 + \mathbf{v}\mathbf{D}_2 | \mathbf{u}\mathbf{D}_3 + \mathbf{v}\mathbf{D}_4) | \mathbf{u} \in U \text{ and } \mathbf{v} \in V\}.$$

It is denoted by  $[U, V] \cdot \mathbf{D}$ .

*Remark 2.* Let  $U$  and  $V$  be codes with generator matrices  $\mathbf{G}_u$  and  $\mathbf{G}_v$ , and parity check matrices  $\mathbf{H}_u$  and  $\mathbf{H}_v$ , respectively. It is a simple exercise to show that

$$\mathbf{G} = \left( \begin{array}{c|c} \mathbf{G}_u\mathbf{D}_1 & \mathbf{G}_u\mathbf{D}_3 \\ \hline \mathbf{G}_v\mathbf{D}_2 & \mathbf{G}_v\mathbf{D}_4 \end{array} \right) \quad \text{and} \quad \mathbf{H} = \left( \begin{array}{c|c} \mathbf{H}_u\mathbf{D}_1 & -\mathbf{H}_u\mathbf{D}_3 \\ \hline \mathbf{H}_v\mathbf{D}_2 & -\mathbf{H}_v\mathbf{D}_4 \end{array} \right)$$

is a generator matrix and a parity check matrix, respectively for  $[U, V] \cdot \mathbf{D}$ .

We consider the matrix-product construction  $[U, V] \cdot \mathbf{D}$  which was already introduced in [5] and rediscovered in [4, 15]. In [4, Theorem3.7] a lower bound for the minimum distance of such code is given when the matrix  $\mathbf{D}$  has a certain property, namely non-singular by columns. In [15] a similar result is proved but makes the hypothesis that  $U$  contains  $V$  (but  $\mathbf{D}$  is arbitrary). Their result does not seem to cover exactly our case, therefore we give a proof below.

**Lemma 10.** *The code  $[U, V] \cdot \mathbf{D}$  has parameters  $[2n, k_u + k_v, d]$  with*

$$\min\{2d_u, d_v\} \leq d \leq \min\{2d_u, 2d_v\}.$$

*Proof.* It is clear that  $\mathcal{C}$  has length  $2n$  and dimension  $k_u + k_v$ .

Now, consider a nonzero codeword  $\mathbf{c} = (\mathbf{u}\mathbf{D}_1 + \mathbf{v}\mathbf{D}_2 | \mathbf{u}\mathbf{D}_3 + \mathbf{v}\mathbf{D}_4) \in \mathcal{C}$ . We distinguish two cases:

- If  $\mathbf{v} = 0$ . Then,  $w_H(\mathbf{c}) = 2w_H(\mathbf{u}) \geq 2d_u$ .
- Otherwise, if  $\mathbf{v} \neq 0$ . By the Triangle Inequality  $w_H(a + b) \geq w_H(a) - w_H(b)$  and the fact that  $\mathbf{D}$  is non singular we have that

$$\begin{aligned} w_H(\mathbf{c}) &= w_H(\mathbf{u}\mathbf{D}_1 + \mathbf{v}\mathbf{D}_2) + w_H(\mathbf{u}\mathbf{D}_3 + \mathbf{v}\mathbf{D}_4) \\ &\geq w_H(\mathbf{u}\mathbf{D}_1 + \mathbf{v}\mathbf{D}_2) + w_H(\mathbf{v}(\mathbf{D}_4 - \mathbf{D}_2\mathbf{D}_1^{-1}\mathbf{D}_3)) - w_H((\mathbf{u}\mathbf{D}_1 + \mathbf{v}\mathbf{D}_2)\mathbf{D}_1^{-1}\mathbf{D}_3) \\ &= w_H(\mathbf{v}) \geq d_v \end{aligned}$$

Thus,  $d \geq \min\{2d_u, d_v\}$ . Moreover, take  $\mathbf{u} \in U$  and  $\mathbf{v} = 0$  with  $w_H(\mathbf{u}) = d_u$ . In such a case  $w_H((\mathbf{u}\mathbf{D}_1 + \mathbf{v}\mathbf{D}_2 \mid \mathbf{u}\mathbf{D}_3 + \mathbf{v}\mathbf{D}_4)) = 2d_u$  and therefore  $d \leq 2d_u$ . The other upper bound follows by choosing  $\mathbf{V} \in V$  and  $\mathbf{u} = 0$  with  $w_H(\mathbf{v}) = d_v$ .  $\square$

Note that the minimum distance of this generalized  $(U \mid U + V)$  construction can supersede the minimum distance of the standard  $(U \mid U + V)$  construction which is equal to  $\min\{2d_u, d_v\}$ .

**Lemma 11.** *The dual code of  $[U, V] \cdot \mathbf{D}$  is the matrix product code  $[U^\perp, V^\perp] \cdot \mathbf{D}'$  with*

$$\mathbf{D}' = \left( \begin{array}{c|c} \mathbf{D}_1 & -\mathbf{D}_3 \\ \hline \mathbf{D}_2 & -\mathbf{D}_4 \end{array} \right)$$

*This code has parameters  $[2n, n - (k_u + k_v), d^\perp]$  with  $\min\{2d_u^\perp, d_v^\perp\} \leq d^\perp \leq \min\{2d_u^\perp, 2dv^\perp\}$ .*

*Proof.* Remark 2 showed that the dual code of  $[U, V] \cdot \mathbf{D}$  is a generalized  $(U \mid U + V)$ -construction of  $U^\perp$  and  $V^\perp$ . Then, the result follows from Lemma 10.  $\square$

These generalized  $(U \mid U + V)$  codes based on RS constituent codes have clearly an efficient decoding which is similar to the  $(U \mid U + V)$ -decoder. There are only a few differences: when we receive a word  $(\mathbf{y}_1, \mathbf{y}_2)$  we just compute the difference  $\mathbf{y}_1\mathbf{D}_3 - \mathbf{y}_2\mathbf{D}_1$  which should be a noisy version of  $\mathbf{v}(\mathbf{D}_2\mathbf{D}_3 - \mathbf{D}_4\mathbf{D}_1)$ . However the error correction capacity is the same as the original  $(U \mid U + V)$  with this kind of decoding algorithm. More precisely, the McEliece scheme we propose is the following

#### Key generation:

- Choose  $U, V$  as RS codes of some length  $n$ .
- Construct a random matrix  $\mathbf{D}$  as described in Definition 2.
- Let  $G$  be a random generator matrix of the code  $\mathcal{C} = [U, V] \cdot \mathbf{D} \cdot \Sigma_{2n}$  where  $\Sigma_{2n}$  is a permutation matrix of size  $2n$  and  $\mathcal{A}_{\mathcal{C}}$  a decoding algorithm for  $\mathcal{C}$  that typically corrects  $t$  errors. It consists in applying  $\Sigma_{2n}^{-1}$  to the received word and then performing the aforementioned generalized  $(U \mid U + V)$ -decoder.

The *public key* and the *private key* are given respectively by:

$$\mathcal{K}_{\text{pub}} = (G, t) \quad \text{and} \quad \mathcal{K}_{\text{secret}} = \mathcal{A}_{\mathcal{C}}$$

**Encryption:**  $\mathbf{y} = \mathbf{m}G + \mathbf{e}$  where  $\mathbf{m}$  is the message and  $\mathbf{e}$  is a random error vector of weight at most  $t$ .

**Decryption:** Use  $\mathcal{K}_{\text{secret}}$  to retrieve  $\mathbf{m}$ .

Thus, by choosing code  $U$  and  $V$  of large enough minimum distance we avoid attacks that try to recover the code structure by looking for low weight codewords either in the code or in its dual. If the minimum distance of the generalized  $(U \mid U + V)$  code is equal to  $2d_u$  such codewords arise as codewords of the form  $(\mathbf{u}\mathbf{D}_1 \mid \mathbf{u}\mathbf{D}_3)\Sigma_{2n}$ . This clearly leaks information about  $\Sigma_{2n}$  and  $\mathbf{D}_1$  and  $\mathbf{D}_3$  if we are able to find such codewords. This is why we want to avoid that such codewords can be easily found.

Note that Wang proposed in [20] a very similar scheme, with the difference that  $U$  was a random code and  $V$  a RS code and that he took only a subcode of the generalized  $(U \mid U + V)$  code namely the code generated by  $\left( \begin{array}{c|c} G_u\mathbf{D}_1 + G_v\mathbf{D}_2 & G_u\mathbf{D}_3 + G_v\mathbf{D}_4 \end{array} \right)$ . The code rate loss implied by this choice results in a significant loss in the key size (since we have to protect ourself against generic decoders for  $t$  errors for a code which is of much smaller dimension). The fact that  $U$  is random in his scheme however is a rather strong argument in favor of its security.

## References

- [1] M. Baldi, M. Bianchi, and F. Chiaraluce. Security and complexity of the McEliece cryptosystem based on QC-LDPC codes. *IET Information Security*, 7(3):212–220, Sept. 2013.
- [2] M. Baldi, M. Bianchi, F. Chiaraluce, J. Rosenthal, and D. Schipani. Enhanced public key security for the McEliece cryptosystem. *J. Cryptology*, 2014.
- [3] A. Bennatan and D. Burshtein. Design and analysis of nonbinary LDPC codes over arbitrary discrete-memoryless channels. *IEEE Trans. Inform. Theory*, 52(2):549–583, Feb. 2006.
- [4] T. Blackmore and H. G. Norton. Matrix-product codes over  $\mathbb{U}_q$ . *Applicable Algebra in Engineering, Communication and Computing*, 12(6):477–500, 2001.
- [5] E. Blokh and H. V. Zyblon. Coding of generalized concatenated codes. *Problems of Information Transmission*, 10:218–222, 1974.
- [6] A. Couvreur, P. Gaborit, V. Gauthier-Umaña, A. Otmani, and J.-P. Tillich. Distinguisher-based attacks on public-key cryptosystems using Reed-Solomon codes. *Des. Codes Cryptogr.*, 73(2):641–666, 2014.
- [7] A. Couvreur, A. Otmani, J. Tillich, and V. Gauthier-Umaña. A polynomial-time attack on the BBCRS scheme. In J. Katz, editor, *Public-Key Cryptography - PKC 2015*, volume 9020 of *Lecture Notes in Comput. Sci.*, pages 175–193. Springer, 2015.
- [8] J.-C. Faugère, A. Otmani, L. Perret, F. de Portzamparc, and J.-P. Tillich. Folding alternant and Goppa Codes with non-trivial automorphism groups. *IEEE Trans. Inform. Theory*, 62(1):184–198, 2016.
- [9] V. Guruswami and A. Rudra. Explicit capacity-achieving list-decodable codes. In *Proceedings of the Thirty-eighth Annual ACM Symposium on Theory of Computing*, STOC ’06, pages 1–10, New York, NY, USA, 2006. ACM.
- [10] V. Guruswami and M. Sudan. Improved decoding of Reed-Solomon and algebraic-geometry codes. *IEEE Trans. Inform. Theory*, 45(6):1757–1767, 1999.
- [11] R. Koetter and A. Vardy. Algebraic soft-decision decoding of reed-solomon codes. *IEEE Trans. Inform. Theory*, 49(11):2809–2825, 2003.
- [12] R. J. McEliece. *A Public-Key System Based on Algebraic Coding Theory*, pages 114–116. Jet Propulsion Lab, 1978.
- [13] R. Misoczki, J.-P. Tillich, N. Sendrier, and P. S. L. M. Barreto. MDPC-McEliece: New McEliece variants from moderate density parity-check codes. In *Proc. IEEE Int. Symposium Inf. Theory - ISIT*, pages 2069–2073, 2013.
- [14] H. Niederreiter. Knapsack-type cryptosystems and algebraic coding theory. *Problems of Control and Information Theory*, 15(2):159–166, 1986.
- [15] F. Özbudak and H. Stichtenoth. Note on niederreiter-xing’s propagation rule for linear codes. *Applicable Algebra in Engineering, Communication and Computing*, 13(1):53–56, 2002.
- [16] F. Parvaresh and A. Vardy. Correcting errors beyond the guruswami-sudan radius in polynomial time. In *Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on*, pages 285–294, 2005.

- [17] P. Shor. Algorithms for quantum computation: Discrete logarithms and factoring. In S. Goldwasser, editor, *FOCS*, pages 124–134, 1994.
- [18] V. M. Sidelnikov and S. Shestakov. On the insecurity of cryptosystems based on generalized Reed-Solomon codes. *Discrete Math. Appl.*, 1(4):439–444, 1992.
- [19] M. Sudan. Decoding of Reed Solomon codes beyond the error-correction bound. *J. Complexity*, 13(1):180–193, 1997.
- [20] Y. Wang. Quantum resistant random linear code based public key encryption scheme RLCE, Dec. 2015.
- [21] C. Wieschebrink. Two NP-complete problems in coding theory with an application in code based cryptography. In *Proc. IEEE Int. Symposium Inf. Theory - ISIT*, pages 1733–1737, 2006.

## A The $(U \mid U + V)$ construction

Along the following two sections and by abuse of notation we will use the symbol  $\mathcal{O}\left(\frac{1}{q}\right)$  not only to describes the big O notation but also as a vector whose  $L$ -infinity norm is behaved like the function  $\frac{1}{q}$  when  $q$  tends to infinity, in other words, a vector whose elements tends to zero as  $q$  goes to infinity . Recall that the infinity norm of a vector  $\mathbf{v} \in \mathbb{F}_q^n$ , denoted  $\|\mathbf{v}\|_\infty$ , is defined as the maximum of the absolute values of its components, i.e.

$$\|\mathbf{v}\|_\infty = \max \{|v(i)| \mid i = 1, \dots, n\}.$$

In this appendix,  $U$  and  $V$  are RS codes and we will use them in a  $(U \mid U + V)$  construction. We start with a  $q$ -ary symmetric channel with error probability  $p$  ( $q$ -SC $_p$ ). On the following we obtain the channel error models for decoding  $U$  and  $V$ . Recall that the reliability matrix for the  $U$ -decoder is  $\Pi_1 = \Pi \times \Pi$  whereas for the  $V$ -decoder it is  $\Pi_2 = \Pi \oplus \Pi$ .

Suppose we transmit the codeword  $\mathbf{u}$  over a noisy channel and we receive the vector

$$\mathbf{y} = (\mathbf{y}_1 \mid \mathbf{y}_2) = \mathbf{u} + (\mathbf{e}_1 \mid \mathbf{e}_2)$$

We begin to observe that in the case of a  $q$ -ary symmetric channel we have only the following possibilities

**Case 1** No error has occurred in position  $i$ :  $e_1(i) = e_2(i) = 0$ .

**Case 2** An error has occurred in position  $i$ :  $e_1(i) \neq 0$  or  $e_2(i) \neq 0$ .

**Case 3** Two errors have occurred in position  $i$ :  $e_1(i) \neq 0$  and  $e_2(i) \neq 0$ .

	Probability of occurrence
<b>Case 1</b> (No errors)	$(1 - p)^2$
<b>Case 2</b> (1 error)	$2p(1 - p)$
<b>Case 3</b> (2 errors)	$p^2$

Table 1: Ways of combining the vectors  $y_1(i)$  and  $y_2(i)$

### A.1 The matrix $\Pi_1 = \Pi \times \Pi$ in the $q$ -ary symmetric channel

**Lemma 12.** *Let  $\pi_U$  be the probability vector corresponding to decoding the code  $U$ . For the channel error model of the code  $U$  we have*

$$\mathbb{E}(\|\pi_U\|^2) = (1-p)^2 + 2p(2-p) \left( \frac{1-p}{2-p} \right)^2 + \mathcal{O}\left(\frac{1}{q}\right) = \frac{(p+2)(p-1)^2}{2-p} + \mathcal{O}\left(\frac{1}{q}\right)$$

*Proof.* We will treat each case as a separate study.

**Case 1** We have (up to permutation)

$$\Pi_{\mathbf{y}_1}^i \times \Pi_{\mathbf{y}_2}^i = \begin{pmatrix} 1-p \\ \frac{p}{q-1} \\ \vdots \\ \frac{p}{q-1} \end{pmatrix} \times \begin{pmatrix} 1-p \\ \frac{p}{q-1} \\ \vdots \\ \frac{p}{q-1} \end{pmatrix} = \begin{pmatrix} \frac{(1-p)^2}{\beta} \\ \frac{p^2}{(q-1)\beta} \\ \vdots \\ \frac{p^2}{(q-1)\beta} \end{pmatrix} = \begin{pmatrix} 1-s \\ \frac{s}{q-1} \\ \vdots \\ \frac{s}{q-1} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \mathcal{O}\left(\frac{1}{q}\right)$$

with  $\beta = (1-p)^2 + (q-1)\frac{p^2}{(q-1)^2}$ . And,  $s = \frac{p^2}{(1-p)^2(q-1)+p^2} = \mathcal{O}\left(\frac{1}{q}\right)$

**Case 2** We have (up to permutation)

$$\Pi_{\mathbf{y}_1}^i \times \Pi_{\mathbf{y}_2}^i = \begin{pmatrix} 1-p \\ \frac{p}{q-1} \\ \frac{p}{q-1} \\ \vdots \\ \frac{p}{q-1} \end{pmatrix} \times \begin{pmatrix} 1-p \\ \frac{p}{q-1} \\ \frac{p}{q-1} \\ \vdots \\ \frac{p}{q-1} \end{pmatrix} = \begin{pmatrix} \frac{(1-p)p}{(q-1)\beta} \\ \frac{(1-p)p}{(q-1)\beta} \\ \frac{p^2}{(q-1)^2\beta} \\ \vdots \\ \frac{p^2}{(q-1)^2\beta} \end{pmatrix} = \begin{pmatrix} \frac{1-p}{2-p} \\ \frac{1-p}{2-p} \\ \frac{p}{(q-1)(2-p)} \\ \vdots \\ \frac{p}{(q-1)(2-p)} \end{pmatrix} + \mathcal{O}\left(\frac{1}{q}\right)$$

with

$$\beta = 2\frac{(1-p)p}{(q-1)} + (q-2)\frac{p^2}{(q-1)^2} = \frac{p(2-p)}{(q-1)} + \mathcal{O}\left(\frac{1}{q}\right)$$

**Case 3** The probability that the same error occurred at  $\mathbf{e}_1(i)$  and  $\mathbf{e}_2(i)$  is assumed to be negligible. Hence (up to permutation),

$$\Pi_{\mathbf{y}_1}^i \times \Pi_{\mathbf{y}_2}^i = \begin{pmatrix} \frac{p}{q-1} \\ 1-p \\ \frac{p}{q-1} \\ \frac{p}{q-1} \\ \vdots \\ \frac{p}{q-1} \end{pmatrix} \times \begin{pmatrix} \frac{p}{q-1} \\ 1-p \\ \frac{p}{q-1} \\ \frac{p}{q-1} \\ \vdots \\ \frac{p}{q-1} \end{pmatrix} = \begin{pmatrix} \frac{1-p}{2-p} \\ \frac{1-p}{2-p} \\ \frac{p}{(q-1)(2-p)} \\ \vdots \\ \frac{p}{(q-1)(2-p)} \end{pmatrix} + \mathcal{O}\left(\frac{1}{q}\right)$$

Once we know all the columns of matrix  $\Pi \times \Pi$ , the result follows easily.  $\square$

Hence, the channel error model of the code  $U$  is represented by the reliability matrix  $\Pi_1 = \Pi \times \Pi$  and the expectation of the  $L_2$  norm  $\|\pi\|^2$  of a column  $\pi$  of  $\Pi_1$  is given by

$$\mathbb{E}(\|\pi_U\|^2) = (1-p)^2 + 2p(2-p) \left( \frac{1-p}{2-p} \right)^2 + \mathcal{O}\left(\frac{1}{q}\right)$$



## A.2 The matrix $\Pi_2 = \Pi \oplus \Pi$ in the $q$ -ary symmetric channel

**Lemma 13.** *Let  $\pi_V$  be the probability vector corresponding to decoding the code  $V$ . The channel error model of the code  $V$  is a  $q$ -SC $_{p'}$  with  $p' = 2p - p^2$  and*

$$\mathbb{E}(\|\pi_V\|^2) = (1 - p')^2 + \mathcal{O}\left(\frac{1}{q}\right) = (1 - p)^4 + \mathcal{O}\left(\frac{1}{q}\right).$$

*Proof.* We will treat each case as a separate study.

**Case 1** No error occurred in position  $i$ , i.e.  $\Pi_{\mathbf{y}_1}^i = \Pi_{\mathbf{y}_2}^i$ . Hence (up to permutation),

$$\Pi_{\mathbf{y}_1}^i \oplus \Pi_{\mathbf{y}_2}^i = \begin{pmatrix} 1-p \\ \frac{p}{q-1} \\ \vdots \\ \frac{p}{q-1} \end{pmatrix} \oplus \begin{pmatrix} 1-p \\ \frac{p}{q-1} \\ \vdots \\ \frac{p}{q-1} \end{pmatrix} = \begin{pmatrix} 1-p' \\ \frac{p'}{q-1} \\ \vdots \\ \frac{p'}{q-1} \end{pmatrix} + \mathcal{O}\left(\frac{1}{q}\right)$$

with  $p' = 2p - p^2$

**Case 2** One error occurred in position  $i$ , in other words, the order of the elements of  $\Pi_{\mathbf{y}_1}^i$  and  $\Pi_{\mathbf{y}_2}^i$  are different. Thus (up to permutation),

$$\Pi_{\mathbf{y}_1}^i \oplus \Pi_{\mathbf{y}_2}^i = \begin{pmatrix} 1-p \\ \frac{p}{q-1} \\ \frac{p}{q-1} \\ \vdots \\ \frac{p}{q-1} \end{pmatrix} \oplus \begin{pmatrix} \frac{p}{q-1} \\ 1-p \\ \frac{p}{q-1} \\ \vdots \\ \frac{p}{q-1} \end{pmatrix} = \begin{pmatrix} \frac{p'}{q-1} \\ 1-p' \\ \frac{p'}{q-1} \\ \vdots \\ \frac{p'}{q-1} \end{pmatrix} + \mathcal{O}\left(\frac{1}{q}\right)$$

with  $p' = 2p - p^2$

**Case 3** We will have two options: either  $\mathbf{e}_1(i) = \mathbf{e}_2(i)$  (similar to **Case 1**, up to permutation) or  $\mathbf{e}_1(i) \neq \mathbf{e}_2(i)$  (similar to **Case 2**, up to permutation).

Thus, the transition matrix  $\Pi \oplus \Pi$  can be represented as a  $q$ -SC $_{p'}$  with  $p' = 2p - p^2$ .  $\square$

## B Recursive application of the $(U \mid U + V)$ construction

In this appendix we study what happens over a  $q$ -ary symmetric channel with error probability  $p$  ( $q$ -SC $_p$ ) if we apply recursively the  $(U \mid U + V)$  construction. That is, we start with a  $(U \mid U + V)$  code, we choose  $U$  to be a  $(U_1 \mid U_1 + V_1)$  code and  $V$  to be a  $(U_2 \mid U_2 + V_2)$  code, where  $U_1$ ,  $U_2$ ,  $V_1$  and  $V_2$  are RS codes over the same alphabet  $\mathbb{F}_q$  and of the same length. In other words, we look for a code of the form

$$(U_1 \mid U_1 + V_1 \mid U_1 + U_2 \mid U_1 + U_2 + V_1 + V_2) = \{(\mathbf{u}_1 \mid \mathbf{u}_1 + \mathbf{v}_1 \mid \mathbf{u}_1 + \mathbf{u}_2 \mid \mathbf{u}_1 + \mathbf{u}_2 + \mathbf{v}_1 + \mathbf{v}_2) : \mathbf{u}_i \in U_i, \mathbf{v}_i \in V_i\}.$$

From Lemma 13 and 12 we will obtain the channel error models for decoding  $U_1$ ,  $V_1$ ,  $U_2$  and  $V_2$  respectively, their reliability matrices are given by  $\Pi_1 \times \Pi_1$ ,  $\Pi_1 \oplus \Pi_1$ ,  $\Pi_2 \times \Pi_2$  and  $\Pi_2 \oplus \Pi_2$  respectively (see Figure 3). We use the previous notation  $p' \stackrel{\text{def}}{=} 2p - p^2$ .

Suppose we transmit the codeword

$$(\mathbf{u}_1 \mid \mathbf{u}_1 + \mathbf{v}_1 \mid \mathbf{u}_1 + \mathbf{u}_2 \mid \mathbf{u}_1 + \mathbf{u}_2 + \mathbf{v}_1 + \mathbf{v}_2)$$

over a noisy channel and we receive the vector

$$\mathbf{y} = (\mathbf{y}_1 \mid \mathbf{y}_2 \mid \mathbf{y}_3 \mid \mathbf{y}_4) = (\mathbf{u}_1 \mid \mathbf{u}_1 + \mathbf{v}_1 \mid \mathbf{u}_1 + \mathbf{u}_2 \mid \mathbf{u}_1 + \mathbf{u}_2 + \mathbf{v}_1 + \mathbf{v}_2) + (\mathbf{e}_1 \mid \mathbf{e}_2 \mid \mathbf{e}_3 \mid \mathbf{e}_4)$$

We begin to observe that in the  $q$ -ary symmetric channel we have only the possibilities given in Table 2

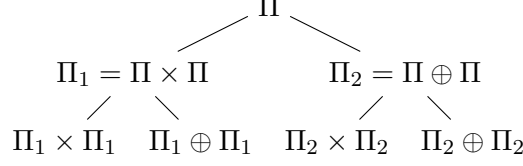


Fig. 3: The different channel error models for  $(U_1 | U_1 + V_1 | U_1 + U_2 | U_1 + U_2 + V_1 + V_2)$

	Result of the combination of ...	Probability of occurrence
<b>Case 4</b> (No errors)	<b>Case 1 and Case 1</b> $e_j(i) = 0$ for all $j \in \{1, 2, 3, 4\}$	$(1 - p)^4$
<b>Case 5</b> (1 error)	<b>Case 1 and Case 2</b> $\exists j_1 \in \{1, 2, 3, 4\}$ s.t. $e_{j_1}(i) \neq 0$ And $e_j(i) = 0$ , otherwise.	$4p(1 - p)^3$
<b>Case 6</b> (2 errors)	<b>Case 1 and Case 3</b> $\exists j_1 \in \{1, 3\}$ s.t. $e_{j_1}(i), e_{j_1+1}(i) \neq 0$ And $e_j(i) = 0$ , otherwise.	$2p^2(1 - p)^2$
<b>Case 7</b> (2 errors)	<b>Case 2 and Case 2</b> $\exists j_1 \in \{1, 2\}$ and $j_2 \in \{3, 4\}$ s.t. $e_{j_1}(i), e_{j_2}(i) \neq 0$ And $e_j(i) = 0$ , otherwise.	$4p^2(1 - p)^2$
<b>Case 8</b> (3 errors)	<b>Case 2 and Case 3</b> $\exists j \in \{1, 2, 3, 4\}$ s.t. $e_{j_1}(i) = 0$ And $e_j(i) \neq 0$ , otherwise.	$4p^3(1 - p)$
<b>Case 9</b> (4 errors)	<b>Case 3 and Case 3</b> $e_j(i) \neq 0$ for all $j \in \{1, 2, 3, 4\}$	$p^4$

Table 2: Ways of combining the vectors  $y_1(i)$ ,  $y_2(i)$ ,  $y_3(i)$  and  $y_4(i)$

### B.1 The matrix $\Pi_2 \oplus \Pi_2$ in the $q$ -ary symmetric channel

**Lemma 14.** Let  $\pi_{V_2}$  be the probability vector corresponding to decoding the code  $V_2$ . The channel error model of the code  $V_2$  is a  $q$ -SC $_{p''}$  with  $p'' = 2p' - p'^2$  and

$$\mathbb{E} \left( \|\pi_{V_2}\|^2 \right) = (1 - p'')^2 + \mathcal{O} \left( \frac{1}{q} \right) = (1 - p)^8 + \mathcal{O} \left( \frac{1}{q} \right)$$

*Proof.* Direct consequence of Lemma 13. □

### B.2 The matrix $\Pi_2 \times \Pi_2$ in the $q$ -ary symmetric channel

**Lemma 15.** Let  $\pi_{U_2}$  be the probability vector corresponding to decoding the code  $U_2$ . We have

$$\mathbb{E} \left( \|\pi_{U_2}\|^2 \right) = \frac{(2 + p')(1 - p')^2}{(2 - p')} + \mathcal{O} \left( \frac{1}{q} \right)$$

*Proof.* Since the transition matrix  $\Pi \oplus \Pi$  can be represented as a  $q$ -SC $_{p'}$  with  $p' = 2p - p^2$ , this case can be treat similar to Lemma 12. □

### B.3 The matrix $\Pi_1 \oplus \Pi_1$ in the $q$ -ary symmetric channel

**Lemma 16.** *Let  $\pi_{V_1}$  be the probability vector corresponding to decoding the code  $V_1$ . We have*

$$\mathbb{E}(\|\pi_{V_1}\|^2) = (1-p)^4 \left( \frac{2+3p+8p^2-4p^3}{2-p} \right) + \mathcal{O}\left(\frac{1}{q}\right)$$

*Proof.* We will treat each case as a separate study.

**Case 4** We have that the column  $(\Pi_{\mathbf{y}_1}^i \times \Pi_{\mathbf{y}_2}^i) \oplus (\Pi_{\mathbf{y}_3}^i \times \Pi_{\mathbf{y}_4}^i)$  is (up to permutation)

$$\left( \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \mathcal{O}\left(\frac{1}{q}\right) \right) \oplus \left( \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \mathcal{O}\left(\frac{1}{q}\right) \right) + \mathcal{O}\left(\frac{1}{q}\right) = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \mathcal{O}\left(\frac{1}{q}\right)$$

**Case 5** We have that the column  $(\Pi_{\mathbf{y}_1}^i \times \Pi_{\mathbf{y}_2}^i) \oplus (\Pi_{\mathbf{y}_3}^i \times \Pi_{\mathbf{y}_4}^i)$  is (up to permutation)

$$\left( \begin{pmatrix} 1+0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \mathcal{O}\left(\frac{1}{q}\right) \right) \oplus \left( \begin{pmatrix} \frac{1-p}{2-p} \\ \frac{1-p}{2-p} \\ \frac{p}{(q-1)(2-p)} \\ \vdots \\ \frac{p}{(q-1)(2-p)} \end{pmatrix} + \mathcal{O}\left(\frac{1}{q}\right) \right) = \begin{pmatrix} \frac{1-p}{2-p} \\ \frac{1-p}{2-p} \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \mathcal{O}\left(\frac{1}{q}\right)$$

**Case 6** Identical result to the above.

**Case 7** We have that the column  $(\Pi_{\mathbf{y}_1}^i \times \Pi_{\mathbf{y}_2}^i) \oplus (\Pi_{\mathbf{y}_3}^i \times \Pi_{\mathbf{y}_4}^i)$  is (up to permutation)

$$\left( \begin{pmatrix} \frac{1-p}{2-p} \\ \frac{1-p}{2-p} \\ \frac{p}{(q-1)(2-p)} \\ \frac{p}{(q-1)(2-p)} \\ \vdots \\ \frac{p}{(q-1)(2-p)} \end{pmatrix} + \mathcal{O}\left(\frac{1}{q}\right) \right) \oplus \left( \begin{pmatrix} \frac{1-p}{2-p} \\ \frac{p}{(q-1)(2-p)} \\ \frac{1-p}{2-p} \\ \frac{p}{(q-1)(2-p)} \\ \vdots \\ \frac{p}{(q-1)(2-p)} \end{pmatrix} + \mathcal{O}\left(\frac{1}{q}\right) \right) = \begin{pmatrix} \left(\frac{1-p}{2-p}\right)^2 \\ \left(\frac{1-p}{2-p}\right)^2 \\ \left(\frac{1-p}{2-p}\right)^2 \\ \left(\frac{1-p}{2-p}\right)^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \mathcal{O}\left(\frac{1}{q}\right)$$

**Case 8** Identical result to the above.

**Case 9** Identical result to the above.

Once we know all the columns of matrix  $\Pi_2 \oplus \Pi_2$ , the result follows easily.  $\square$

### B.4 The matrix $\Pi_1 \times \Pi_1$ in the $q$ -ary symmetric channel

**Lemma 17.** *Let  $\pi_{U_1}$  be the probability vector corresponding to decoding the code  $U_1$ . We have*

$$\mathbb{E}(\|\pi_{U_1}\|^2) = \frac{(5p^3 - 6p^2 - 5p - 4)(1-p)^2}{4-3p} + \mathcal{O}\left(\frac{1}{q}\right)$$

*Proof.* We study separately three different cases:

- **Case 4, Case 5 and Case 6.** In all these cases the columns of the transition matrix have the same form (up to permutation).

$$\Pi_{\mathbf{y}_1}^i \times \Pi_{\mathbf{y}_2}^i \times \Pi_{\mathbf{y}_3}^i \times \Pi_{\mathbf{y}_4}^i = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \mathcal{O}\left(\frac{1}{q}\right)$$

- **Case 7** We have that the column  $\Pi_{\mathbf{y}_1}^i \times \Pi_{\mathbf{y}_2}^i \times \Pi_{\mathbf{y}_3}^i \times \Pi_{\mathbf{y}_4}^i$  is (up to permutation)

$$\begin{aligned} & \left( \begin{pmatrix} \frac{1-p}{2-p} \\ \frac{1-p}{2-p} \\ \frac{p}{(q-1)(2-p)} \\ \frac{p}{(q-1)(2-p)} \\ \vdots \\ \frac{p}{(q-1)(2-p)} \end{pmatrix} + \mathcal{O}\left(\frac{1}{q}\right) \right) \times \left( \begin{pmatrix} \frac{1-p}{2-p} \\ \frac{p}{(q-1)(2-p)} \\ \frac{1-p}{2-p} \\ \frac{p}{(q-1)(2-p)} \\ \vdots \\ \frac{p}{(q-1)(2-p)} \end{pmatrix} + \mathcal{O}\left(\frac{1}{q}\right) \right) = \begin{pmatrix} \left(\frac{1-p}{2-p}\right)^2 \frac{1}{\alpha} \\ \frac{(1-p)p}{(2-p)^2(q-1)} \frac{1}{\alpha} \\ \frac{(1-p)p}{(2-p)^2(q-1)} \frac{1}{\alpha} \\ \frac{p^2}{(q-1)^2(2-p)^2} \frac{1}{\alpha} \\ \vdots \\ \frac{p^2}{(q-1)^2(2-p)^2} \frac{1}{\alpha} \end{pmatrix} + \mathcal{O}\left(\frac{1}{q}\right) \\ & = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \mathcal{O}\left(\frac{1}{q}\right) \end{aligned}$$

$$\text{with } \alpha = \left(\frac{1-p}{2-p}\right)^2 + 2\frac{(1-p)p}{(2-p)^2(q-1)} + (q-3)\frac{p^2}{(q-1)^2(2-p)^2} = \left(\frac{1-p}{2-p}\right)^2 + \mathcal{O}\left(\frac{1}{q}\right)$$

- **Case 8** We have that the column  $\Pi_{\mathbf{y}_1}^i \times \Pi_{\mathbf{y}_2}^i \times \Pi_{\mathbf{y}_3}^i \times \Pi_{\mathbf{y}_4}^i$  is (up to permutation)

$$\begin{aligned} & \left( \begin{pmatrix} \frac{1-p}{2-p} \\ \frac{1-p}{2-p} \\ \frac{p}{(q-1)(2-p)} \\ \frac{p}{(q-1)(2-p)} \\ \vdots \\ \frac{p}{(q-1)(2-p)} \end{pmatrix} + \mathcal{O}\left(\frac{1}{q}\right) \right) \times \left( \begin{pmatrix} \frac{p}{(q-1)(2-p)} \\ \frac{p}{(q-1)(2-p)} \\ \frac{1-p}{2-p} \\ \frac{1-p}{2-p} \\ \vdots \\ \frac{p}{(q-1)(2-p)} \end{pmatrix} + \mathcal{O}\left(\frac{1}{q}\right) \right) = \begin{pmatrix} \frac{(1-p)p}{(2-p)^2(q-1)} \frac{1}{\alpha} \\ \frac{(1-p)p}{(2-p)^2(q-1)} \frac{1}{\alpha} \\ \frac{(1-p)p}{(2-p)^2(q-1)} \frac{1}{\alpha} \\ \frac{(1-p)p}{(2-p)^2(q-1)} \frac{1}{\alpha} \\ \frac{p^2}{(q-1)^2(2-p)^2} \frac{1}{\alpha} \\ \vdots \\ \frac{p^2}{(q-1)^2(2-p)^2} \frac{1}{\alpha} \end{pmatrix} + \mathcal{O}\left(\frac{1}{q}\right) \\ & = \begin{pmatrix} \frac{1-p}{4-3p} \\ \frac{1-p}{4-3p} \\ \frac{1-p}{4-3p} \\ \frac{1-p}{4-3p} \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \mathcal{O}\left(\frac{1}{q}\right) \end{aligned}$$

with

$$\alpha = 4\frac{(1-p)p}{(2-p)^2(q-1)} + (q-4)\frac{p^2}{(q-1)^2(2-p)^2} = \frac{p(4-3p)}{(2-p)^2(q-1)} + \mathcal{O}\left(\frac{1}{q}\right)$$

- **Case 9** Identical result to the above.

Once we know all the columns of matrix  $\Pi_2 \times \Pi_2$ , the result follows easily.

□